LAW OFFICES
McGuireWoods LLP
1750 TYSONS BOULEVARD, SUITE 1800
MCLEAN, VIRGINIA 22102

APPLICATION FOR UNITED STATES LETTERS PATENT

Applicants:	Huajie Chen, Dureseti Chidambarrao and Omer H. Dokumaci
For:	HIGH PERFORMANCE STRESS-ENHANCED MOSFETs USING Si:C AND SiGe EPITAXIAL SOURCE/DRAIN AND METHOD OF MANUFACTURE
Docket No.:	FIS920030241US1

HIGH PERFORMANCE STRESS-ENHANCED MOSFETs USING Si:C and SiGe EPITAXIAL SOURCE/DRAIN AND METHOD OF MANUFACTURE

DESCRIPTION

BACKGROUND OF THE INVENTION

Field of the Invention

The invention generally relates to a semiconductor device and method of manufacture and, more particularly, to a semiconductor device and method of manufacture which imposes tensile and compressive stresses in the device during device fabrication.

Background Description

Mechanical stresses within a semiconductor device substrate can modulate device performance. That is, stresses within a semiconductor device are known to enhance semiconductor device characteristics. Thus, to improve the characteristics of a semiconductor device, tensile and/or compressive stresses are created in the channel of the n-type devices (e.g., NFETs) and/or p-type devices (e.g., PFETs). However, the same stress component, either tensile stress or compressive stress, discriminatively affects the characteristics of an n-type device and a p-type device.

For example, it has been known that a device exhibits better performance characteristics when formed on a silicon layer (or cap) that is epitaxially grown on another epitaxially grown SiGe layer that has relaxed on top of the silicon substrate. In

this system, the silicon cap experiences biaxial tensile strain. When epitaxially grown on silicon, an unrelaxed SiGe layer will have a lattice constant that conforms to that of the silicon substrate. Upon relaxation (through a high temperature process for example) the SiGe lattice constant approaches that of its intrinsic lattice constant which is larger than that of silicon. A fully relaxed SiGe layer has a lattice constant close to that of its intrinsic value. When the silicon layer is epitaxially grown thereon, the silicon layer conforms to the larger lattice constant of the relaxed SiGe layer and this applies physical biaxial stress (e.g., expansion) to the silicon layer being formed thereon. This physical stress applied to the silicon layer is beneficial to the devices (e.g., CMOS devices) formed thereon because the expanded silicon layer increases N type device performance and a higher Ge concentration in the SiGe layer improves P type device performances.

Relaxation in SiGe on silicon substrates occurs through the formation of misfit dislocations. For a perfectly relaxed substrate, one can envision a grid of misfit dislocations equally spaced that relieve the stress. The misfit dislocations facilitate the lattice constant in the SiGe layer to seek its intrinsic value by providing extra half-planes of silicon in the substrate. The mismatch strain across the SiGe/silicon interface is then accommodated and the SiGe lattice constant is allowed to get larger.

However, the problem with this conventional approach is that it requires a multi-layered SiGe buffer layer that is very thick (e.g., a thickness of approximately 5000 Å to 15000 Å) to achieve misfit dislocations on its surface portion while avoiding threading dislocations between the SiGe layer and the silicon substrate layer, thereby achieving a relaxed SiGe structure on the surface of the multi-layered SiGe layer. Also, this approach

significantly increases manufacturing time and costs. Further, the thick graded SiGe buffer layer cannot be easily applicable to silicon-on-insulator substrate (SOI). This is because for silicon-on-insulator the silicon thickness has to be below 1500 Å for the benefits of SOI to be valid. The SiGe buffered layer structure is too thick.

Another problem is that misfit dislocations formed between the SiGe layer and the silicon epitaxial layer are random and highly non-uniform and cannot be easily controlled due to heterogeneous nucleation that cannot be easily controlled. Also, misfit dislocation densities are significantly different from one place to another. Thus, the physical stress derived from the non-uniform misfit dislocations are apt to be also highly non-uniform in the silicon epitaxial layer, and this non-uniform stress causes non-uniform benefits for performance with larger variability. Further at those locations where misfit density is high, the defects degrade device performances through shorting device terminals and through other significant leakage mechanisms.

It is also known to grow Si:C epitaxially on Si where it is inherently tensile. A 1% C content in a Si:C/Si material stack can cause tensile stress levels in the Si:C on the order of 500MPa. In comparison, in the SiGe/Si system about 6% Ge is needed to causes a 500MPa compression. This 1% level of C can be incorporated into Si during epitaxial growth as shown in Ernst et al., VLSI Symp., 2002, p92. In Ernst et al., Si/Si:C/Si layered channels for nFETs are shown. However in the Ernst et al. structure, the Si:C is provided in the traditional strained Si approach as a stack of layers in the channel. Thus, in the Ernst et al. structure, the Si:C is used as part of the channel, itself. The problem

with this approach is that the mobility is not enhanced, but retarded depending on the C content from scattering.

SUMMARY OF THE INVENTION

In a first aspect of the invention, a method of manufacturing a semiconductor structure includes forming a p-type field-effect-transistor (pFET) channel and a negative field-effect-transistor (nFET) channel in a substrate. A pFET stack and an nFET stack are formed on the substrate associated with the respective channels. A first layer of material is provided at source/drain regions associated with the pFET stack. The first layer of material has a lattice constant different than a base lattice constant of the substrate to create a compressive state within the pFET channel. A second layer of material is provided at the source/drain regions associated with the nFET stack. The second layer of material has a lattice constant different than the base lattice constant of the substrate to create a tensile state at the nFET channel.

In another aspect of the invention, a method of manufacturing a semiconductor structure is provided which includes forming a p-type field-effect-transistor (pFET) channel and a negative field-effect-transistor (nFET) channel in a substrate. A pFET structure and an nFET structure are formed on the substrate associated with the pFET channel and the nFET channel, respectively. The regions of the pFET structure and the nFET structure are etched to a predetermined depth. A first material with a lattice constant different than the base lattice constant of the layer is provided in the etched regions of the pFET structure to provide a compressive stress in the pFET channel. A second material with a lattice constant different than the base lattice constant of the layer

is provided in the etched regions of the nFET structure to provide a tensile stress in the nFET channel. Doping source and drain regions of the nFET and pFET structures is provided.

In yet another aspect of the invention, a semiconductor structure includes a semiconductor substrate, and a pFET and nFET are formed in respective channels in the substrate. A first layer of material in source and drain regions of the pFET channel has a lattice constant different than the lattice constant of the substrate. A second layer of material in source and drain regions of the nFET channel has a lattice constant different than the lattice constant of the substrate.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1a through 1e represent a fabrication process to form a device in accordance with the invention;

Figure 2 illustrates the locations of the stresses in an pFET device according to the invention; and

Figure 3 illustrates the locations of the stresses in an nFET device according to the invention.

DETAILED DESCRIPTION OF EMBODIMENTS OF THE INVENTION

The invention is directed to a semiconductor device and method of manufacture which provides tensile stresses near the nFET channel and compressive stresses near the

pFET channel of CMOS devices. In one embodiment of the invention, the longitudinal tensile stresses are brought very close to the nFET channel while at the same time compressive stresses are brought very close to the pFET channel. Additionally, in the invention, a process and structure is provided to integrate both SiGe and Si:C materials into CMOS technology.

By way of example, a highly tensile Si:C film is provided (e.g., embedded) in the silicon substrate in the source/drain (S/D) regions to longitudinally apply tension on the nFET structure in the channel under the gate region thereof. Similarly, a highly compressive SiGe film is provided (e.g., embedded) in the silicon substrate in the S/D regions to longitudinally apply compression on the pFET in the channel under the gate region thereof. Similar to the SiGe layer, the Si:C layer is relatively thin (below its critical thickness) and is not relaxed. The transistor channel region of the nFET is strained by the stress from the Si:C layer, while the channel region of the pFET is provided with compressive stress from the SiGe.

Since the SiGe layer is embedded in the S/D region of the pFET, low resistance silicide can still be formed. Interestingly, embedded (e.g., sub-surface or coplanar-to-surface) Si:C films can put larger stresses than the above-surface Si:C counterparts because of the film surface not being free. In this invention, the different thicknesses and protrusions of the Si:C, whether embedded or coplanar with the surface or raised, are contemplated. It should be understood that by adjusting the concentrations of the Ge content in the SiGe layer, it is possible to adjust the compressive stress in the pFET channel. Similarly, it is possible to adjust the tensile stress in the nFET channel by

adjusting the concentrations of C in the Si:C layer. This is due to the lattice constant of such materials.

In the invention, it is shown that Si:C has a correct stress, that it includes the correct amount of C content, and that it can be grown epitaxially and selectively. Also, in the invention, the Si:C is not used as a directly built up stacked layer under the channel but as a replacement material for nFET S/D regions that are in tension and therefore imposing tension in the channel region. This Si:C film therefore applies tensile stress to nFET channels while the SiGe applies compressive stress to the pFET channels.

Figures 1a through 1e represent a fabrication process to form a device in accordance with the invention. In Figure 1a, a silicon-on-insulator (SOI) 20 or the like is provided. The layer 20 is patterned to form shallow trench isolation features (STI) 25 using standard techniques of pad oxidation, pad nitride deposition, lithography based patterning, reactive ion etching (RIE) of the stack consisting of nitride, oxide, and silicon down to the buried oxide, edge oxidation, liner deposition, fill deposition, and chemical mechanical polish, for example. The STI formation process is well known in the art. The pad nitride is then stripped. The gate stacks comprising, for example, gate dielectric and poly silicon are formed on the structure to form the pFETs and the nFETs, in any well known method. TEOS caps 43 and 44 are formed with the pPETs and nFETs in a known manner.

Still referring to Figure 1a, spacers are formed on the pFET and nFET stacks, respectively, using any well known process. By way of example, spacers 38 are formed

on the side walls of the pFET stack 40a and spacers 42 are formed on the sidewalls of the nFET stack 45a. The spacers may be oxide or nitride spacers, for example.

In Figure 1b, a thin liner 50 is blanket deposited over the structure including the pFET stack, nFET stack and S/D regions thereof, for example. The thin liner 50 is, in one embodiment, a Si_3N_4 liner or any of a nitride or oxide based material depending on the material of a hard mask. The thickness range of the thin liner is approximately 5 to 20 nm. The thin liner 50 may act as a protection layer. A hard mask 51 is then formed over the nFET stack 45a and S/D regions thereof.

Regions about the pFET stack 40a are etched to the liner 50. The liner 50 is then etched and S/D regions 52 are formed (etched) adjacent the stack 40a. The depth of S/D regions 52 is about 20 to 100 nm, depending on the thickness of the SOI layer. A highly compressive selective epitaxial SiGe layer 60 is grown in the regions 52 of the pFET stack 40a as shown in Figure 1c, fully filling the S/D etched regions 52. The SiGe layer 60 may be grown to a thickness of about 10 to 100 nm thick, although other thicknesses are also contemplated by the invention. In one implementation, the SiGe layer is grown to a thickness above the surface of the gate oxide. The hard mask and remaining portions of the liner are removed using any well known process such as, for example, wet chemicals. In processing steps, dopants are ion implanted to form the S/D regions about the pFET stack 40a prior to removing the hard mask.

Standing alone, the SiGe normally has a larger lattice constant than the SOI. That is, the lattice constant of the SiGe material does not match the lattice constant of the Si.

However, in the structure of the invention, due to the growth of the SiGe layer, the lattice structure of the SiGe layer will tend to match the lattice structure of the underlying Si. This results in the SiGe layer and the channel regions adjacent or next to the SiGe being under compression. In one embodiment, the Ge content of the SiGe layer may be greater than 0% and upwards ratio to the Si content.

Still referring to the processing of the invention, in Figure 1d, a thin liner 50 is again blanket deposited over the structure including the nFET, pFET and S/D regions thereof, for example. The thin liner 50 is, in one embodiment, a Si_3N_4 liner or any of a nitride or oxide material depending on the material of a hard mask. The thickness range of the thin liner is approximately 5 to 20 nm. The thin liner 50 may act as a protection layer.

A mask 51 is then formed over the pFET stack 40a, and the regions about the nFET stack 45a are etched to the liner 50. The liner is then etched and S/D regions 54 are formed (etched) adjacent the stack 45a. The depth of the S/D regions 54 is about 20 to 100 nm, depending on the thickness of the SOI layer. Any well known process may be used to etch the regions 54.

A highly tensile selective epitaxial Si:C layer is grown to a thickness of about 10 to 100 nm thick in the regions 54 of the nFET stack 45a as shown in Figure 1e. It should be understood that the Si:C layer 62 may be epitaxial grown to other thicknesses, as contemplated by the invention. In one embodiment, the C content may be from greater than 0% to 4% in ratio to the Si content. The resist and remaining portions of the thin

liner are removed using any well known process such as, for example using wet chemicals.

Standing alone, Si:C would normally have a smaller lattice constant than the underlying Si. That is, the lattice constant of the Si:C material does not match the lattice constant of the Si. However, in the structure of the invention, due to the growth of the Si:C layer within the S/D regions of the nFET stack 45a, the lattice structure of the Si:C layer will tend to match the lattice structure of the underlying Si. This results in the Si:C layer and the channel regions adjacent or next to the Si:C being under a tensile stress.

In one embodiment, the Si:C and/or the SiGe layer may be embedded within the device by, example, performing a RIE, wet etching or other processes or combinations thereof for processing optimization to recess such layers. Then, Si is selectively grown over the regions 52 and 54 using any well known process. Source and drain implanting may then be performed using any well process. Further processing can be performed to build the devices and interconnects.

As now should be understood, both the Si:C and SiGe layers may be embedded within the device or can be coplanar or raised from the device. In one embodiment, the Si:C and SiGe films are deposited to 10 to 100 nm thick, which provides a more cost-effective way to add stress (compressive or tensile) to the MOSFETs. In the raised embodiment, the Si:C and SiGe layers may be raised above the surface of the device to about 50 nm. It should be recognized that other thicknesses are also contemplated by the invention.

Also, it is also possible to in situ doping the SiGe with P type doping and Si:C with n type doping to form your source and drain regions of the pFET and nFET, respectively.

It should also be understood by those of ordinary skill in the art that the process steps of Figures 1d and 1e may be equally be performed prior to the process steps shown in Figures 1b and 1c. Also, further processing steps such as, for example, standard ion implantation can be performed to form the S/D regions of the pFETs and nFETs. The formation of the S/D regions, through the ion implantation, is self aligning due to the gate oxide in the nFET and pFET regions acting as a mask during this process.

Figure 2 illustrates the locations of the stresses in an nFET device according to the invention. As shown in Figure 2, compressive stresses are present in just below the pFET with a region of unrelaxed SiGe. More specifically, in the structure of the invention, the lattice structure of the SiGe layer matches the lattice structure of the underlying Si insulation layer. This results in the SiGe layer and the surrounding areas being under a compressive stress.

Figure 3 illustrates the locations of the stresses in a nFET device according to the invention. As shown in Figure 3, tensile stresses are present in the channel of the nFET. More specifically, in the structure of the invention, the lattice structure of the Si:C layer will match the lattice structure of the underlying Si insulation layer 20 to form a tensile stress component in the nFET channel.

While the invention has been described in terms of embodiments, those skilled in the art will recognize that the invention can be practiced with modifications and in the spirit and scope of the appended claims. For example, the invention can be readily applicable to bulk substrates.